

13/07/2020

Aplicacions de l'aprenentatge profund' (*Deep Learning*) per a la millora genètica dels poliploides



Un article recentment publicat per autors del Centre de Recerca en Agrigenòmica (CRAG) i de la Universitat de Florida mostra com els mètodes computacionals moderns poden ajudar en la millora genètica d'espècies vegetals amb més de dos jocs de cromosomes.

La capacitat predictiva de les tècniques d'aprenentatge profund s'ha avaluat en nabius i maduixes (Crèdit: Muffet / CC BY)

El desenvolupament de noves varietats de plantes sol consistir en un procés de prova i error en el qual es creuen línies 'elit' consolidades i s'avalua el rendiment de la seva descendència. Amb el temps, alguns d'aquests descendents substitueixen les línies "elit" quan les superen en almenys alguns dels caràcters d'interès, com ara la resistència a les malalties o el gust. Aquest és un procés continu, però, per desgràcia, lent. Per exemple, el temps per desenvolupar una nova varietat genètica de maduixa és de més de vuit anys.

Molts trets d'interès en plantes són complexos

Per aquesta raó, els "milloradors" s'han esforçat a accelerar aquest procés utilitzant tecnologies genòmiques modernes. Una possibilitat és dur a terme proves genètiques per identificar els encreuaments i individus més favorables, cosa que requereix conèixer els gens i les mutacions

responsables del caràcter. Malauradament, i contrari al que se sol pensar, els gens que determinen els trets d'interès econòmic només es coneixen parcialment; de fet, fins ara s'han descobert relativament poques mutacions causals. A més, l'expressió d'un caràcter, com pot ser el gust, depèn no només dels gens, sinó també de les condicions ambientals en què creixen les plantes. Per exemple, la mateixa varietat pot créixer molt bé si la irrigació és suficient, però alhora ser molt sensible a la sequera. El personal responsable de la millora de plantes defineix aquests trets com a 'complexos', ja que depenen de l'entorn i de molts gens que només estan caracteritzats parcialment.

Què es pot fer en aquests casos? Un cop més, els mètodes moleculars poden ajudar-nos, però utilitzant una aproximació complementària anomenada 'predicció genòmica'. Aquest procediment consisteix a utilitzar tots els marcadors genètics disponibles, incloent-hi candidats que no s'han validat prèviament, per 'predir' els resultats dels encreuaments entre diferents varietats vegetals. Això es fa típicament usant variants del conegut mètode de regressió lineal.

Com pot ajudar l'aprenentatge profund

Existeixen nombrosos mètodes de predicció genòmica, però la majoria consideren una relació relativament simple entre els marcadors genètics i el tret d'interès. No obstant això, recentment s'han desenvolupat nous mètodes de predicció denominats 'aprenentatge profund', *deep learning* en anglès. L'aprenentatge profund comprèn un conjunt d'algoritmes que s'inspiren en el funcionament del cervell humà i que divideixen el procediment de càlcul en petites unitats anomenades 'neurones'. Aquests mètodes són molt populars avui dia i tenen nombroses aplicacions, que van de la traducció automàtica a l'anàlisi de vídeo i so.

L'interès de l'aprenentatge profund en la genòmica rau en la gran flexibilitat pel que fa a la relació que assumeix entre els marcadors i els caràcters d'interès. Això és important en espècies com la maduixa perquè són poliploides –tenen més de dues còpies del mateix cromosoma–, i és precisament en aquest tipus d'espècies que esperem que les interaccions entre els gens siguin més importants del que és habitual. L'aprenentatge profund pot ser una eina prometedora per a la predicció genòmica en aquest escenari.

El treball de Zingaretti et al., realitzat pel Centre de Recerca en Agrigenòmica (CRAG) en col·laboració amb la Universitat de Florida, aborda precisament aquesta qüestió. En aquest estudi mostrem, en una de les primeres aplicacions de l'aprenentatge profund a la predicció genòmica, que aquest mètode pot ser realment eficaç en presència d'interaccions entre gens, és a dir, quan un caràcter no es pot predir simplement tenint en compte els gens de manera individual. Ho demostrem amb dues espècies d'interès hortícola: la maduixa, que és octoploide, i el nabiu, tetraploide. També proporcionem programes informàtics que poden ser utilitzats pel personal investigador per aplicar l'aprenentatge profund a la predicció genòmica. Aquest treball és d'especial rellevància per a la indústria espanyola, ja que, segons dades del 2018, Espanya és el primer productor europeu de maduixes i el sisè en l'àmbit mundial.

Miguel Pérez-Enciso

Profesor d'investigació ICREA a la UAB.

Centre de Recerca en Agrigenòmica (CRAG) CSIC-IRTA-UAB-UB.

miguel.perez@cragenomica.es

Referències

L.M. Zingaretti, S.A. Gezan, L.F. Ferrão, L.F. Osorio, A. Monfort, P.R. Muñoz, V.M. Whitaker, M. Pérez-Enciso. 2020. **Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species**. *Frontiers in Plant Science* doi.org/10.3389/fpls.2020.00025.

Pérez-Enciso M, Zingaretti LM. 2019. **A Guide on Deep Learning for Complex Trait Genomic Prediction**. *Genes* doi.org/10.3390/genes10070553.

[View low-bandwidth version](#)